Transportation Research Part C 163 (2024) 104659

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Unravelling uncertainty in trajectory prediction using a non-parametric approach

Guopeng Li [a,*], Zirui Li [b], Victor L. Knoop [a], Hans van Lint [a]

[a] Transport and Planning, Civil Engineering and Geosciences, Delft University of Technology, Mekelweg 5, 2628 CD Delft, Netherlands
[b] The School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Keywords:
Trajectory prediction
Uncertainty quantification
Microscopic traffic modelling
Causal inference

ABSTRACT

Predicting the trajectories of road agents is fundamental for self-driving cars. Trajectory prediction contains many sources of uncertainty in data and modelling. A thorough understanding of this uncertainty is crucial in a safety-critical task like auto-piloting a vehicle. In practice, it is necessary to distinguish between the uncertainty caused by partial observability of all factors that may affect a driver's near-future decisions, the so-called *aleatoric uncertainty*, and the uncertainty of deploying a model in new scenarios that are possibly not present in the training set, the so-called *epistemic uncertainty*. They reflect the trade-off between data collection and model improvement In this paper, we propose a new framework to systematically quantify both sources of uncertainty. Specifically, to approximate the spatial distribution of an agent's future position, we propose a 2D histogram-based deep learning model combined with deep ensemble techniques for measuring aleatoric and epistemic uncertainty by entropy-based quantities. The proposed Uncertainty Quantification Network (UQnet) employs a causal part to enhance its generalizability so rare driving behaviours can be effectively identified. Experiments on the INTERACTION dataset show that UQnet is able to give more robust predictions in generalizability tests compared to the correlation-based models. Further analysis presents that high aleatoric uncertainty cases are mainly caused by heterogeneous driving behaviours and unknown intended directions. Based on this aleatoric uncertainty component, we estimate the lower bounds of mean-square-error and final-displacement-error as indicators for the predictability of trajectories. Furthermore, the analysis of epistemic uncertainty illustrates that domain knowledge of speed-dependent driving behaviour is essential for adapting a model from low-speed to high-speed situations. Our paper contributes to motion forecasting with a new framework, that recasts the problem of accuracy improvement in a way that focuses on differentiating between unpredictable components and rare cases for which more and different data should be collected.

1. Introduction

1.1. Background

Predicting human-driven vehicles' intentions, behaviours, and trajectories is an essentially important topic for engineers and researchers in many domains (Lefèvre et al., 2014). For example, in the context of autonomous driving, motion forecasting is indispensable for developing safe and smooth self-driving systems (Yoon et al., 2019). Accurate and reliable trajectory prediction is

* Corresponding author.
E-mail addresses: g.li-5@tudelft.nl (G. Li), z.li@bit.edu.cn (Z. Li), v.l.knoop@tudelft.nl (V.L. Knoop), j.w.c.vanlint@tudelft.nl (H. van Lint).

https://doi.org/10.1016/j.trc.2024.104659
Received 7 October 2022; Received in revised form 11 June 2023; Accepted 5 May 2024
Available online 12 May 2024

critical for motion planning (Wang et al., 2021). Trajectory prediction is also very important for multi-scale traffic modelling. Many phenomena in traffic flow originate from microscopic driving behaviours. For instance, heterogeneous car-following behaviours may cause road capacity drop (Yuan et al., 2018) and improper lane-changing around an on-ramp can lead to traffic congestion (Daamen et al., 2010; Leclercq et al., 2016). Building a reactive microscopic traffic simulator from real-world data requires high-quality trajectory prediction in the forward simulation (Bergamini et al., 2021). Many applications, such as dynamic traffic signal control (Chen et al., 2020) and collaborative platoon cruising (Hallé and Chaib-draa, 2005), can benefit from such a reactive simulator, and thus improve the macroscopic efficiency of traffic networks.

Recently, emerging Artificial Intelligence (AI) techniques and public trajectory datasets, such as Waymo (Sun et al., 2020), nuScenes (Caesar et al., 2020), Argoverse (Chang et al., 2019), etc. together stimulate the fast development of trajectory forecasting (Rudenko et al., 2020). In the literature, numerous Deep Neural Networks (DNN) have been proposed to continuously improve prediction performance. Significant progress has been made by applying AI to trajectory-based vehicle motion prediction. We refer the readers to Huang et al. (2022) for a comprehensive survey.

Although trajectory data are fundamental for motion forecasting, they clearly do not disclose the full complexity of driving behaviours. First, observed trajectories are just the final results of the underlying interactions between vehicles and the environment (including other vehicles). Many important factors, such as driving styles, vehicle characteristics and all the factors that influence these (e.g. turn signals), are not observable using trajectories alone. Second, there is always a non-zero probability of encountering rare behaviours (e.g. a risky cut-in) or circumstances (e.g. a combination of high demand and signal malfunctioning) which may lead to different resulting interactions and thus different trajectories than present in the training data set. The model faces a much higher risk of getting it wrong in such a situation. Either way, *uncertainty* is unavoidable in trajectory-based motion prediction. Ignoring uncertainty will make the self-driving system unaware of potential danger and thus lead to accidents. Uncertainty also poses the boundary of one model's capability and depicts the unpredictable part of the future given the recent observed data. Before quantifying uncertainty, it is necessary to clarify different types of uncertainty and their specific roles in motion prediction.

1.2. Uncertainty in motion forecasting

From a modelling and practical perspective, the total predictive uncertainty can be categorized into two types, **aleatoric** uncertainty and **epistemic** uncertainty (Der Kiureghian and Ditlevsen, 2009).

Aleatoric uncertainty represents the inherent randomness in the collected data and the underlying process itself. For motion prediction, aleatoric uncertainty has two major sources. The first one, which we call endogenous *interaction uncertainty*, is due to the fact that human drivers do not share all the information needed to predict the intended future behaviours of all other (relevant) drivers. As a result, they can at best make an informed guess about the driving strategies of their surrounding drivers before the interaction starts, and use this guess to, for example, cooperate or not (Lutteken et al., 2016). The consequence of this interaction uncertainty is that future trajectories may be completely different even with similar starting conditions. In a lane-changing case, for example, the following vehicle on the adjacent lane may yield to the cutting-in vehicle, or drive more aggressively and force the other driver to abandon the lane-change (Wang et al., 2015). The second source of aleatoric uncertainty is due to the discrepancy between the information perceived by the so-called demonstrators (in this case the human drivers) and the imitators (sensors that observe these demonstrators), a problem which is also referred to as *sensor shift* (Etesami and Geiger, 2020). Restricted by perception methods, imitators do not perceive as much or even the same information as the demonstrators. For example, if trajectories are collected by drones, important stimuli such as traffic lights or turn signals are not observable due to the bird-eye-view position. Without these stimuli, the uncertainty around the predicted future state is much larger, particularly at decision points on the road (e.g. intersections). Also, datasets collected by sensors installed on vehicles, such as cameras, Lidar, Radar, etc, may be "blind" to relevant signals that can potentially influence the demonstrators' decisions, for example, sound, glare, in-vehicle information, etc.

In summary, aleatoric uncertainty is the result of limitations in observability, either from the perspective of demonstrators (the agents whose path we aim to predict) or the imitators (the sensors partially measuring this path). We cannot distinguish between these components from data only, and the combined aleatoric uncertainty is thus *irreducible*, regardless of the size of the data set. The total aleatoric uncertainty can be understood as an upper bound of predictive accuracy for all models using the same type(s) of data. Aleatoric uncertainty is therefore also called *data uncertainty*. If a model's performance is close enough to this limit, investing in building newer and larger models cannot bring significant accuracy improvement. The types of data must be diversified.

Epistemic uncertainty, or alternatively, *knowledge uncertainty*, represents the uncertainty that in principle could be reduced to zero with the data available. Nonzero epistemic uncertainty is due to the "rareness" of the prevailing situation. In the case of trajectory prediction, it measures whether enough similar samples have already been seen to support reliable trajectory prediction. Under many circumstances, road users behave in similar ways (Makansi et al., 2021), which implies a prediction of this behaviour can be reliable most of the time. However, rare cases typically pertain to unsafe and high-risk situations. Finding out these "corner cases" by the prediction model itself is the key to building an "honest and trustworthy" self-driving system that can clearly tell what it does not know. Quantifying the "rareness" of these samples is also important for anomaly detection (Laxhammar and Falkman, 2013), continuous learning (Ebrahimi et al., 2019), and evaluating the generalizability of a model.

Although aleatoric and epistemic uncertainty quantification (UQ) has already been widely studied, especially in combination with deep learning models (Abdar et al., 2021 provides a comprehensive review), applying it in motion forecasting has radically different requirements.

First, a driver's intention is rigorously restricted by *arbitrary* layouts of road networks. Therefore, the spatial distribution (of probable trajectories) cannot always be approximated by simple priors (e.g. a 2D or mixture Gaussian) (Gilles et al., 2021).

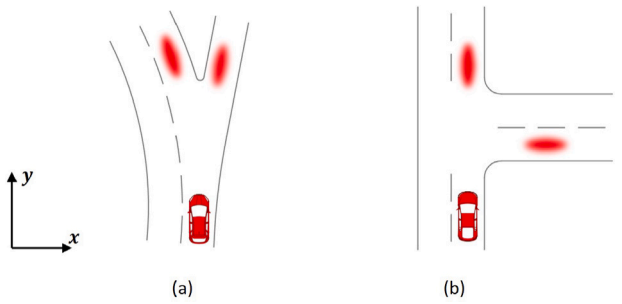


Fig. 1. The spatial probability distribution of the target vehicle's future position in two different scenarios. They have different covariance matrices but the same differential entropy.

Second, the longitudinal and lateral components of 2D coordinates can be strongly correlated, yielding inflated or deflated uncertainty. If we use the covariance matrix as in classical UQ approaches, those non-diagonal elements are not intuitively explainable. Further, if the prediction has strong multi-modality (multiple local maxima), those diagonal elements are not meaningful either, regardless of the determinant of the covariance matrix. Fig. 1 illustrates the problem. The left and right scenarios in Fig. 1 represent different layouts (Fork versus T-junction) but exactly the same decision problem (left or right?) and thus the same degree of uncertainty in motion prediction. However, the co-variance in x and y ($\mathrm{Var}(x)$, $\mathrm{Var}(y)$), and the determinant of the covariance matrix in the left case (the fork) are all significantly smaller than the right case (the T-junction). We thus need to use a more reasonable *scalar* uncertainty metric that can exclude this artefact due to road layouts, in which the angle of two diverging or merging roads may inflate or deflate the estimated uncertainty.

Third, we argue that epistemic uncertainty estimation is closely related to *domain adaptation* in motion prediction. The aim of quantifying epistemic uncertainty is detecting *rare driving behaviours* (represented by trajectories). However, trajectories are strongly *correlated* with the topology of lane networks. If the model cannot adapt itself to those unseen test lane networks by learning the correct *causal effects* of surrounding road agents, many normal driving behaviours in new scenarios will be incorrectly recognized as rare samples. This will lead to significant degradation in epistemic uncertainty estimation.

In summary, the 3 points above must be addressed: a suitable uncertainty metric; a non-parametric representation method; and a generalizable model. Then we formulate the research questions central to this study:

How predictable are vehicular trajectories, especially the final point? What aspects must be considered to generalize prediction models and uncertainty quantification methods to new cases?

1.3. Contributions and outline

Inspired by recent works on motion prediction (Gu et al., 2021; Gilles et al., 2021) and UQ method (Malinin and Gales, 2018), we propose a non-parametric approach to estimate both aleatoric and epistemic uncertainty in human drivers' trajectory forecasting. Instead of using a set of parameters to assimilate the closed-form distribution of future positions, our model directly learns a mesh-grid 2D histogram (a heatmap) to approximate any distribution. This heatmap-based model is combined with deep ensemble techniques to quantify predictive uncertainty. Scalar entropy-based quantities are used as uncertainty metrics. The conditional differential entropy and the mutual information represent aleatoric and epistemic uncertainty respectively. To make the estimate of epistemic uncertainty more reliable, a regularization net is added to the predictor to suppress (control for) spurious correlations. Experiments on the INTERACTION dataset (Zhan et al., 2019) show that the proposed UQnet method indeed has better generalizability. Based on the estimated uncertainty, we further analyse the predictability of vehicle trajectories and illustrate how speed is a key factor in model generalization.

The major contributions of this paper are summarized below:

- Propose a deep-ensemble-based non-parametric approach for quantifying both aleatoric and epistemic uncertainty measured by entropy quantities in single-agent trajectory prediction.
- Induce a causal regularization to enhance the generalizability and make the estimated epistemic uncertainty more reliable.
- Use quantified uncertainty for analysing the predictability of trajectory and detecting out-of-distribution cases. The results further give insights into the role of speed in model generalization and vehicle interaction.

The paper is organized as follows. We first briefly overview the relevant works in Section 2. Next, we introduce the proposed method and the used model UQnet in Sections 3 and 4. Section 5 shows the experimental results and the corresponding analysis. Finally, we draw our conclusions and give several research directions in the last section.

2. Overview

In this section, we will present a short overview of related studies in the literature and bring out the distinctiveness of the proposed method.

There are different approaches to predicting vehicle trajectories. One popular approach is to apply simple physics-model-based methods in combination with (e.g. Kalman) filtering (Prevost et al., 2007), in which typical assumptions such as constant yaw rate and/or acceleration are made to simplify the problem with prior co-variance structures that model unobservable deviations (Houenou et al., 2013; Ammoun and Nashashibi, 2009). These models describe the projected movements in explanatory terms but typically suffer from limitations in prediction horizons and accuracy (Lefèvre et al., 2014).

On the other hand, recently-popular data-driven models, especially deep-learning models, directly assimilate the collected data and show better performances in many scenarios. For example, Alahi et al. (2016) proposed to model interactions between pedestrians with social pooling in a "black-box" way. Here "social" represents the common scene rules and the influence of neighbour agents. Further studies improve the social mechanism by inducing generative adversarial networks (Gupta et al., 2018) or considering multi-agent dynamic features (Zhao et al., 2019). Some studies abstract agents as nodes, treat pairwise influences as edges and use graph neural networks to model the interaction, e.g. Vemula et al. (2018), Ma et al. (2019). Huang et al. (2019) further seek "post-hoc" interpretation from learnt graph attention weights, saying higher attention means larger influence. *Trajectron++* (Salzmann et al., 2020) developed a modular and graph-structured recurrent model as the encoding channel to generate multi-modal predictions from incorporated agent dynamics. VectorNet (Gao et al., 2020) proposed an alternative vector-based lightweight representation to reduce model complexity. It has been widely used in many following works (e.g. Liang et al., 2020).

However, the models above highly rely on independent identical distribution (i.i.d.) assumption. Compared to physical models, correlation-based DNNs are fragile when being deployed in new environments because of causal confusion (De Haan et al., 2019), which means the model learns spurious correlations (overfits the problem). Recently some studies tried to improve DNN's generalizability by inducing more robust causal relationships. Chen et al. (2021) proposed to use counterfactual analysis to alleviate the spurious (specious) correlation of environmental bias. The plug-in module consistently improves the performance of baseline models. Hu et al. (2021) constructed a structural causal model (SCM) to learn invariant features across different scenarios, the so-called causal-based time series domain generalization (CTSDG) model. In Liu et al. (2022), the input is decomposed into invariant variables, style confounders, and spurious features in hidden space. By training the model to suppress spurious features, the robustness was significantly improved. Kumor et al. (2021) came up with a theoretical criterion that determines the feasibility of learning a demonstrator's trajectory under sensor shift from the perspective of causal models. In brief, combining DNNs with a causal model is expected to increase both the robustness and the transparency of data-driven trajectory prediction models.

In addition to limits in generalizability, uncertainty is another important topic in motion forecasting, but it has not drawn as much attention as developing prediction models. Arnez et al. (2020) reviewed different UQ approaches that potentially can be used for autonomous vehicle applications. For example, Makansi et al. (2019) proposed a sampling-fitting two-stage strategy to learn the mixture Gaussian distribution of a vehicle's future position, which can naturally represent aleatoric uncertainty. In Pang et al. (2021), a Bayesian neural network is used to quantify the total uncertainty brought by weather for trajectory forecasting. Djuric et al. (2020) directly give estimates of both aleatoric and epistemic uncertainties measured by marginal variance for each prediction step in vehicle trajectory forecasting. The model was also deployed in a real-world autonomous vehicle test. Tang et al. (2022) further consider the prediction uncertainty for safer decision-making and motion planning in high-risk scenarios. The authors assume that the predicted position obeys a Gaussian distribution and use Deep Ensembles to explicitly quantify both aleatoric and epistemic uncertainty measured by the covariance matrix. The studies above pioneered this domain. They are all parametric approaches, which means the prior form of spatial distribution must be closed-form (such as Gaussian, Laplace, or Mixture Density). However, this assumption does not always hold. For example, vehicles tend to drive along the lane centrelines so the form of its spatial distribution highly depends on the road layouts and lane connectivity. Recently, heatmap-based models pave a path to non-parametric UQ. Based on the vectorized representation, DenseTNT (Gu et al., 2021) converts the 2D regression problem to a classification task and directly learns the probability that the target vehicle will appear in each small mesh unit. This method is anchor-free and it can approximate any 2D distribution. GoHome (Gilles et al., 2021), similarly, generates lane-centralized heatmap from high-definite maps and trajectories. The authors explicitly point out that this approach can be used to quantify aleatoric and epistemic uncertainty. In summary, we believe that there are already enough tools towards more reasonable UQ in motion prediction.

From the overview above, we find there are 3 key gaps in the literature. First, most papers use a variance/covariance matrix to quantify uncertainty, which is not reasonable for a 2D distribution with multiple local maxima (multi-modal). Second, domain adaptation and uncertainty quantification are regarded as two separate topics. We argue that they are highly entangled, especially for quantifying epistemic uncertainty. Third, the predictability of trajectories hidden in the heatmap-based representation is barely discussed. Most studies focus on accuracy improvement instead of the limit of this improvement. Our method tries to overcome the first drawback by using entropy-based metrics for heatmap outputs and to bridge the second gap by inducing causation in deep-learning models. Then the quantified aleatoric and epistemic uncertainty will be used to answer the predictability question.

3. Method

This section describes the proposed method in detail. We will sequentially introduce the 2D histogram and the Deep-Ensemble-based uncertainty quantification method and how to combine this technique with Granger causality.

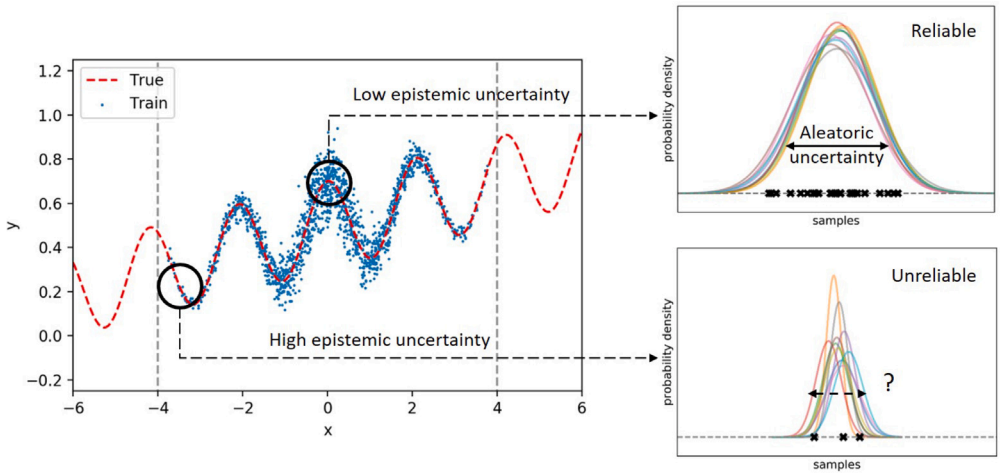


Fig. 2. A 1D example of aleatoric and epistemic uncertainty. Here we learn $y = f(x)$ from noisy data. Both the magnitude of noise and the number of samples are higher in the middle but decay with $|x|$.

3.1. Uncertainty quantification

Suppose that the input and the output of a model are two random variables, respectively noted as $\boldsymbol{X}$ and $\boldsymbol{Y}$. The training dataset contains N collected input–output samples, noted as $\mathcal{D}_N = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^N$. $\mathcal{D}_N$ is used to train a prediction model with inner random variables $\boldsymbol{\Theta} \sim p_\Theta(\boldsymbol{\theta})$, noted as $p_Y(\boldsymbol{y}) = \mathcal{M}(\boldsymbol{x}, \boldsymbol{\theta})$. Now given a test input $\boldsymbol{x}^*$, we aim to quantify the uncertainty of the output distribution $p_Y(\boldsymbol{y}|\boldsymbol{X} = \boldsymbol{x}^*)$ (briefly denoted as $p(\boldsymbol{y}^*)$ and the specific random variable is noted as $\boldsymbol{Y}^*$ from now on). In this study we use *differential entropy* $H(\boldsymbol{Y}^*)$, which is a scalar metric, to represent the total prediction uncertainty. It is defined by:

$$H(\boldsymbol{Y}^*) = -\int_{\mathcal{Y}} p(\boldsymbol{y}^*) \ln p(\boldsymbol{y}^*) d\boldsymbol{y}^* \tag{1}$$

For the multi-modal distributions shown in Fig. 1, the two local maxima are separated by road boundaries so apparently both cases have the same entropy. This metric can exclude the influence of road layouts. The unit of differential entropy is "nats". Amini et al. (2020) derive that the total entropy can be decomposed into the following terms:

$$H(\boldsymbol{Y}^*) = \underbrace{\mathbb{E}_{p_\Theta(\boldsymbol{\theta})}[H(\boldsymbol{Y}^*|\Theta = \boldsymbol{\theta})]}_{\text{aleatoric}} + \underbrace{\mathbb{E}_{p_\Theta(\boldsymbol{\theta})}[D_{KL}(p(\boldsymbol{y}^*|\Theta = \boldsymbol{\theta}) \parallel p(\boldsymbol{y}^*))]}_{\text{epistemic}} \tag{2}$$

Here $p(\boldsymbol{y}^*)$ is the posterior distribution marginalized by $\boldsymbol{\theta}$, given as follows:

$$p(\boldsymbol{y}^*) = \mathbb{E}_{p_\Theta(\boldsymbol{\theta})}[p(\boldsymbol{y}^*|\Theta = \boldsymbol{\theta})] \tag{3}$$

And $D_{KL}(p \parallel q)$ is the *Kullback–Leibler* divergence that measures the directed "distance" from distribution p to q, which is non-negative. The definition is:

$$D_{KL}(p(\boldsymbol{y}^*|\Theta = \boldsymbol{\theta}) \parallel p(\boldsymbol{y}^*)) = \int_{\mathcal{Y}} p(\boldsymbol{y}^*|\Theta = \boldsymbol{\theta}) \ln \frac{p(\boldsymbol{y}^*|\Theta = \boldsymbol{\theta})}{p(\boldsymbol{y}^*)} d\boldsymbol{y}^* \tag{4}$$

So (2) can be interpreted as follows. The first aleatoric term is the *conditional entropy* that measures the average entropy (uncertainty) across an ensemble of distributions. The second epistemic term is the *mutual information* that measures the average distance from each distribution to the average (posterior) distribution, which reflects how diverse the ensemble of distributions is. During inference, if the model has already seen similar inputs enough times in training, then the output distributions will be consistent for different model parameters $\boldsymbol{\theta}$. If the input is a rare or new case, different $\boldsymbol{\theta}$ will give diverse distributions because it contributes little or even zero to the training loss. This can be illustrated by the simple 1D example in Fig. 2. Different models will give consistent predictions at $x = 0$ but diverse outputs at $x = -3.5$. Here we also point out one special property: Different from covariance measures, the epistemic term in (2) is **scale-independent**. It is independent of the mean or variance of each distribution, which makes it naturally convenient to represent "rareness" objectively.

The discussion above shows the principle of UQ, that is learning $p_\Theta(\boldsymbol{\theta})$ that gives consistent predictions when enough samples are provided but diverse outputs for rare samples. The existing methods in the literature include Bayesian neural networks, Monte-Carlo dropout (Kendall and Gal, 2017), and Deep Ensemble (DE) (Lakshminarayanan et al., 2017), etc. Among these UQ methods, the deep ensemble is still the most robust approach and it is state-of-the-art in many uncertainty quantification tasks (e.g. monocular depth estimation (Poggi et al., 2020)). An ensemble of randomly-initialized models are trained independently. During inference, the input is passed into the trained models in parallel to get their corresponding output distributions. Fort et al. (2019) interpret the

advantage of DE from the perspective of loss landscape. The authors show that deep ensembles can explore different local minima of the loss while other methods usually fall into only one. In this study, we also choose the deep ensemble strategy.

Most studies in the literature assume that the prior form of the output distribution can be described by a set of parameters $\boldsymbol{\gamma} = \mathcal{M}(\boldsymbol{x}^*, \boldsymbol{\theta})$, such as mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ for Gaussian (Makansi et al., 2019), or concentration $\boldsymbol{\alpha}$ for Dirichlet distribution, etc. However, the output distribution can be much more complex due to arbitrary road layouts. To alleviate this restriction, we directly approximate a 2D spatial distribution by a mesh-grid heatmap $\hat{\boldsymbol{Y}}_{h\times w} = \mathcal{M}(\boldsymbol{x}^*, \boldsymbol{\theta})$. The value at a pixel $\hat{Y}_{i,j}$ represents the probability that the vehicle will be present within that specific square. So the regression problem is converted to a dense multi-class classification problem. An ensemble of N models give N 2D distributions $\{\hat{\boldsymbol{Y}}_n\}_{n=1}^N$ for one input $\boldsymbol{x}^*$. Then the posterior distribution $p(\boldsymbol{y}^*)$ can be approximated by the element-wise average distribution:

$$p(\boldsymbol{y}^*) \approx \boldsymbol{Y}_m = \frac{1}{N}\sum_{n=1}^{N} \hat{\boldsymbol{Y}}_n \tag{5}$$

And the conditional differential entropy of each heatmap can be approximated by (here x and y are coordinates):

$$H_n = -\int_{\mathcal{X},\mathcal{Y}} \hat{Y}_n(x,y) \ln \hat{Y}_n(x,y) dx dy \tag{6}$$

The KL divergence between each distribution to the posterior distribution can be approximated by:

$$D_{KL}(\hat{\boldsymbol{Y}}_n \| \boldsymbol{Y}_m) = \int_{\mathcal{S}} \hat{Y}_n(s) \ln \frac{\hat{Y}_n(s)}{Y_m(s)} ds \tag{7}$$

Where $\mathcal{S}$ is the non-zero support set of $\boldsymbol{Y}_m$ (to avoid 0 division). Therefore, according to (2), the aleatoric term can be approximated by the average of conditional differential entropy:

$$\mathbb{E}_{p_\Theta(\theta)}[H(\boldsymbol{Y}^*|\Theta = \theta)] \approx \frac{1}{N}\sum_{n=1}^{N} H_n \tag{8}$$

And the epistemic term can be approximated by the average of KL divergence:

$$\mathbb{E}_{p_\Theta(\theta)}[D_{KL}(p(\boldsymbol{y}^*|\Theta = \theta) \parallel p(\boldsymbol{y}^*))] \approx \frac{1}{N}\sum_{n=1}^{N} D_{KL}(\boldsymbol{Y}_n \| \boldsymbol{Y}_m) \tag{9}$$

In summary, (8) and (9) give the estimates of aleatoric and epistemic uncertainty from deep ensembles respectively. For the 2D integrals in (6) and (7), we use *Simpson's rule* to calculate them numerically. We refer the readers to Cruz-Uribe and Neugebauer (2002) for the approximation error bound of this method. Besides the numerical error of integral, the accuracy of epistemic term estimation (9) largely depends on the size of deep ensembles N. If N is not large enough, the absolute error can be high due to the bad approximation of Eq. (5). However, if we are only interested in the relative rareness of samples, this approach works well with a smaller ensemble size.

3.2. Causal regularization

The UQ method introduced above completely ignores what assumptions are used in the model. Considering $\boldsymbol{X}$ and $\boldsymbol{Y}$ as two *correlated* random variables works well when the identical independent distribution assumption holds but becomes fragile for out-of-distribution (OOD) samples. This is especially problematic for model generalization and epistemic uncertainty quantification. Specific to microscopic traffic modelling, the input $\boldsymbol{X}$ can be decomposed into three variables:

(1) $\boldsymbol{M}$: Maps information, including the geometries of lanes and their connectivity, traffic rules, etc.
(2) $\boldsymbol{E}$: The observed trajectory of the target vehicle.
(3) $\boldsymbol{S}$: The observed trajectories of the surrounding agents (including vehicles and pedestrians).

One of the most significant spurious correlations is between $\boldsymbol{M}$ and $\boldsymbol{S}$ (shown in Fig. 3(a)). For example, in the training scenario shown in Fig. 3(b), a driver simply follows the surrounding vehicles OR follows the lane centreline works perfectly fine most of the time. This may cause severe over-fitting to the scenario biases. When deploying the trained model in a new merging where two roads are not parallel, this correlation does not exist anymore. For the specific case on the right figure, we observed that different deep learning models will give two separate groups of predictions: follow the other vehicles (off-road black star) or follow the lane (in-road green star). It means that the DE will unreasonably identify this case as "rare". In Bahari et al. (2022), the authors also show that most state-of-the-art motion forecasting models can be confused by new road layouts and give off-road outputs.

Inducing *causation* is one plausible solution. Because $\boldsymbol{M}$, $\boldsymbol{E}$, and $\boldsymbol{S}$ are given for each specific prediction, we formulate the relationship as *Granger causality* (Granger, 1980). We say a variable "Granger-causes" $\boldsymbol{Y}$ if involving this variable can reduce the uncertainty of $\boldsymbol{Y}$ ("narrow down" the expected occupied area). Maps information and the past trajectory of the target vehicle must be given otherwise we can predict nothing. So here we aim to learn the correct causal effect of $\boldsymbol{S}$. We induce an additional regularization net which only takes $\boldsymbol{M}$ and $\boldsymbol{E}$ as inputs, noted as $\widetilde{\boldsymbol{Y}} = \mathcal{M}_{\text{rg}}(\boldsymbol{M}, \boldsymbol{E}|\boldsymbol{M})$. The output of the normal predictor net (including $\boldsymbol{S}$) is noted as $\hat{\boldsymbol{Y}}$. The causal effect of inducing $\boldsymbol{S}$ should make the output distribution (1) equally or more concentrated and (2) stay within the drivable area given by excluding surrounding agents. Mathematically we can quantify the causal effect by:

$$\arg\min L_1(\hat{\boldsymbol{Y}}_{01}, \min(\hat{\boldsymbol{Y}}_{01}, \widetilde{\boldsymbol{Y}}_{01})) \tag{10}$$

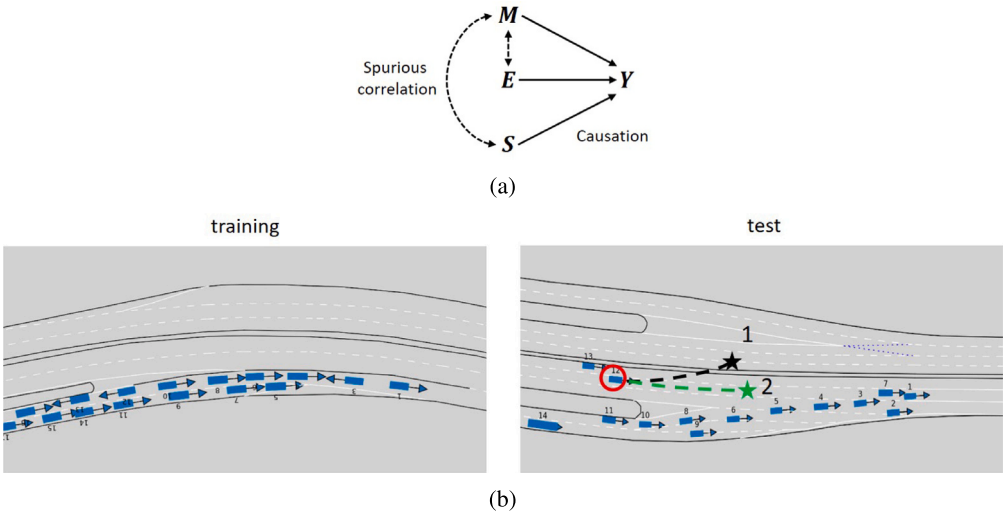


Fig. 3. (a) The causal model of trajectory prediction. Solid lines are causation and dotted lines are spurious correlations. (b) In the training set scenario, the driving direction and the average traffic flow are highly correlated. But the correlation does not hold in the test scenario so a correlation-based model may fail.

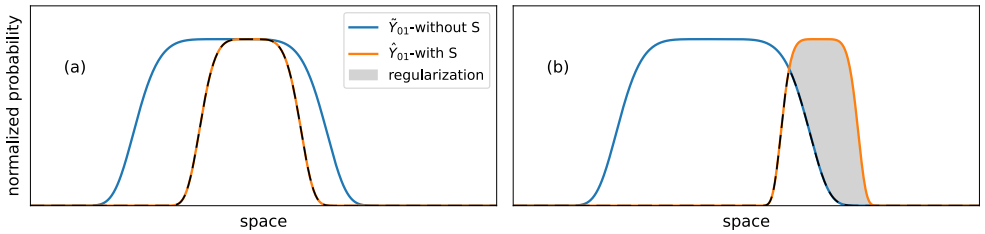


Fig. 4. 1D example of the causal regularization. (a) is a normal case and (b) breaks the correlation of $\boldsymbol{S}$ and $\boldsymbol{M}$.

L_1 is L1 norm, and the subscript "01" means re-normalizing the initial distribution linearly between 0 and 1. min here is the element-wise minimum.

During training, only some driving behaviours, like lane-changing, U-turning etc., may break the correlation between $\boldsymbol{S}$ and $\boldsymbol{M}$ and allow learning the effect of Eq. (10) (If these driving behaviours do not exist at all in the training set, theoretically, correlation and causation are indistinguishable by any means). So, the contributions of these samples must be larger in the total loss. Assume that the loss function between the ground-truth and the prediction is $L(\hat{\boldsymbol{Y}}, \boldsymbol{Y})$. We apply a prediction-dependent weight for each specific sample. The weight increases with the effect given in (10). So the total loss function is constructed as follows:

$$L_{\text{total}}(\widetilde{\boldsymbol{Y}}, \hat{\boldsymbol{Y}}, \boldsymbol{Y}) = (1 + \underbrace{\lambda L_1(\hat{\boldsymbol{Y}}_{01}, \min(\hat{\boldsymbol{Y}}_{01}, \widetilde{\boldsymbol{Y}}_{01}))}_{\text{regularization}}) \underbrace{L(\hat{\boldsymbol{Y}}, \boldsymbol{Y})}_{\text{predictor loss}} + \underbrace{L(\widetilde{\boldsymbol{Y}}, \boldsymbol{Y})}_{\text{regularization loss}} \tag{11}$$

Now we intuitively explain the regularization term by using the 1-D example in Fig. 4. The regularization net is trained independently. Without information about surrounding agents (but the label indeed considers the surrounding vehicles), it will give an equally or more uncertain output $\widetilde{\boldsymbol{Y}}$ that covers all possibilities (the blue distribution). If the normal prediction that considers surrounding vehicles $\hat{\boldsymbol{Y}}$ is within this area, then the regularization term (10) is 0 (see case-(a) in Fig. 4), which means all variables are correlated in this sample. Nothing can be done. When the correlation between $\boldsymbol{S}$ and $\boldsymbol{M}$ is broken in a training sample, and $\hat{\boldsymbol{Y}}$ is partly or completely out of $\widetilde{\boldsymbol{Y}}$ (as shown in case-(b) in Fig. 4), the regularization term (shaded area) will assign a higher weight to $L(\hat{\boldsymbol{Y}}, \boldsymbol{Y})$. In this way, the model is forced to focus on learning the correct effect of $\boldsymbol{S}$ so the two groups of predictions shown in Fig. 3 will reduce to one. Here λ is a hyper-parameter. A lower percentage of correlation-broken samples in the training set requires a larger λ. In Appendix A, a Bayesian explanation is also presented. We refer the readers there for details.

In summary, the proposed method employs heatmap-based deep ensembles to quantify predictive uncertainty. A regularization net and a causation-based regularization term are added to enhance the model's adaptability. Next, we will present the structure of the proposed model.

4. UQnet model

Training an ensemble of deep-learning-based motion forecasting models is time-consuming. Running the inference also requires higher memory and a longer time. Therefore, fast and lightweight models are preferred. In this study, we use the representation

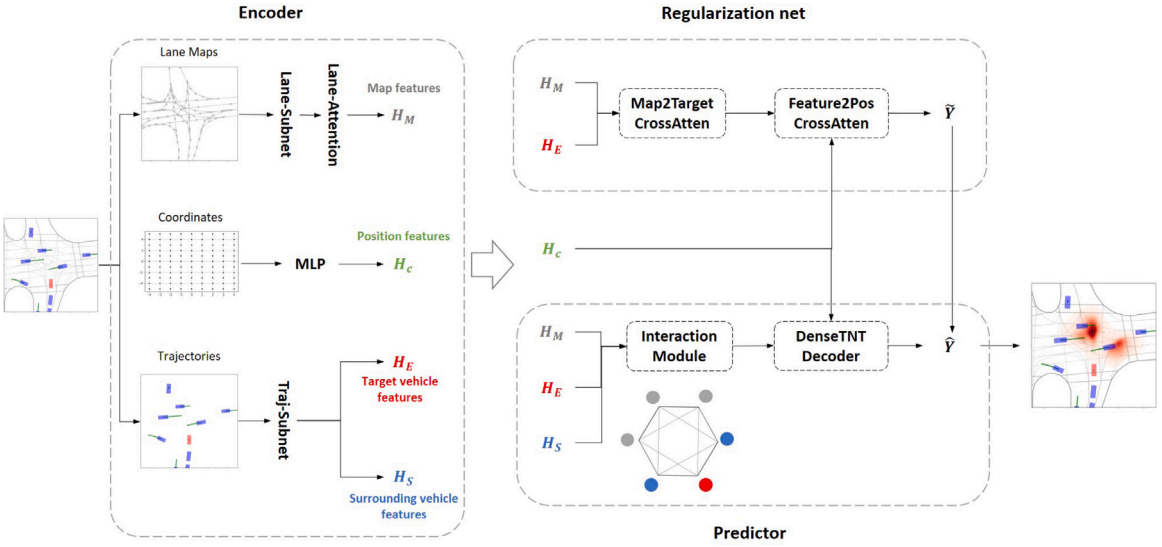


Fig. 5. Model structure of UQnet.

proposed in VectorNet (Gao et al., 2020). VectorNet abstracts map elements (such as lane centrelines, crosswalks, etc.) and agents' trajectories into splines. Each spline is represented by a series of end-to-head connected vectors. Compared with other representation methods, such as rasterized images, this vector-based representation significantly reduces the input size but preserves the most important factors in motion forecasting.

The backbone of our predictor is constructed based on DenseTNT (Gu et al., 2021). Its structure is shown in Fig. 5. The predictor and the counterfactual net share the same sub graph neural networks (encoder) to extract features from each spline separately. Then we added a laneGCN-like (Liang et al., 2020) graph attention module (*Lane-Attention*) to explicitly learn the connectivity of roads. The basic idea is that upstream/downstream, left/right adjacent lanes have different influences on the central lane. We use an attention-based module instead of the parameter-fixed laneGCN module proposed in Liang et al. (2020) so the influence of one lane on another depends on their own features as well. Meanwhile, a multi-layer perceptron (MLP) is used to convert each coordinate (x_t, y_t) to hidden representation $\boldsymbol{H}_C$. The extracted hidden representations of maps elements ($\boldsymbol{H}_M$), coordinates ($\boldsymbol{H}_C$), and trajectories ($\boldsymbol{H}_E$ for target vehicle, $\boldsymbol{H}_S$ for surrounding agents) are passed into the predictor and the regularization net.

The regularization net decoder was constructed based on the proposed causal relationship. A cross-attention layer (*Map2Target*) passes map information to the target vehicle. Then another cross attention layer (*Feature2Pos*) generates the probability density of each location from the concatenated output of *Map2Target* layer. In the predictor decoder, stacked multi-head self-attention graph layers are employed to learn the interactions among these splines. At least two layers are needed to learn their state-dependent dynamic relationships. Then the same decoder used in DenseTNT (Gu et al., 2021) generates the prediction heatmap $\hat{\boldsymbol{Y}}$. For more details, we refer to Appendix B and the open source code.[1] We name our model *Uncertainty Quantification networks* (**UQnet**).

UQnet uses *Focal loss* (Lin et al., 2017) as $L()$ to measure the error between the ground-truth and the predicted distribution. It is defined as follows:

$$L_{\text{fl}}(\hat{\boldsymbol{Y}}, \boldsymbol{Y}) = -\frac{1}{P}\sum_{p}(Y_p - \hat{Y}_p)^2 f(Y_p, \hat{Y}_p) \tag{12}$$

$$f(Y_p, \hat{Y}_p) = \begin{cases} \ln \hat{Y}_p & \text{if } Y_p = 1 \\ (1 - Y_p)^4 \ln(1 - \hat{Y}_p) & \text{else} \end{cases} \tag{13}$$

Focal loss can well address imbalanced samples. The ground-truth label is constructed by the objection detection technique proposed in Zhou et al. (2019) and also used in GoHome (Gilles et al., 2021). For example, if the ground-truth location is (x_g, y_g). We add an extra Gaussian noise ϵ. Its mean is the ground-truth location and the covariance matrix has the form $\sigma_\epsilon^2 \boldsymbol{I}$. For every pixel $Y_{i,j}$ of the heatmap with centre coordinate (x_c, y_c), its probability is determined as follows:

$$Y_{i,j} = \frac{1}{2\pi\sigma_\epsilon^2}\exp\left[-\frac{(x_c - x_g)^2 + (y_c - y_g)^2}{2\sigma_\epsilon^2}\right] \tag{14}$$

This method increases the number of non-zero pixels in the heatmap and thus can significantly accelerate the training process. However, it also induces extra errors. In this study, we choose $\sigma_\epsilon = 0.7\,\text{m}$ so the Gaussian roughly covers the average length of a

[1] https://github.com/RomainLITUD/UQnet-arxiv/

car. It also means that we manually increase the mean-square-error (MSE) by $0.49\,\mathrm{m}^2$ and the lower bound of differential entropy is the entropy of this white noise, around 2.12. We need to correct this in uncertainty estimation. Replacing L in (11) by (12), we have the total loss function to train UQnet. The value of λ depends on the size of the output heatmap and its spatial resolution. In inference, the regularization net can be removed and we only keep the predictor. The lanescore module in DenseTNT decoder is also used to increase the convergence speed, but we do not consider lanescore in inference or uncertainty quantification.

5. Evaluation

In this section, the proposed method will be evaluated on the open INTERACTION (Zhan et al., 2019) dataset. INTERACTION collects trajectories of road agents by drones in diverse urban traffic scenarios and high-definite maps are also provided. In the INTERPRET single-agent track prediction challenge, all agents' trajectories in the past 1 s are provided to predict the track of the target vehicle in the next 3 s. The data providers split all these cases into three groups. The training set contains 47584 cases in 12 scenarios. The validation set has 11794 cases from the same scenarios. The test set has 22644 cases and around 30% of them are collected from new scenarios to measure the generalizability of the model. For fairness, labels of the test set are not provided.

For each case, all trajectories of agents and lane splines are centred to the target vehicle's current position and re-oriented according to its driving direction (y-axis points at the yaw angle). Each lane's centreline is evenly split into 5 head-to-end vectors along the driving direction. Each centreline vector has 8 features, $\boldsymbol{f}_i = [\boldsymbol{x}_s, \boldsymbol{x}_e, j, c, l, w]$. The first two are positions of start and end points. The integer j is the order of the vector. We also incorporate lane-level features. $c \in \{0, 1\}$ represents whether this lane intersects with another non-connected lane. w is 1/2 width of the lane, ranging from 1.5 m to 4.5 m and l is the total length of the lane. For trajectory vectors we use the similar representation $\boldsymbol{f}_a = [\boldsymbol{x}_s, \boldsymbol{x}_e, \boldsymbol{v}, \text{agent\_type}, t]$. $\boldsymbol{v}$ is the average speed between 2 sequential timestamps. The integer agent type can be a vehicle (1) or pedestrian/cyclist (−1) and t is the timestamp. From the description above we can construct the input for UQnet. For example, if the current case has n_m lanes and n_a agents, Then the maps input is $\boldsymbol{M} \in \mathrm{R}^{n_m \times 5 \times 8}$ and the trajectories input is $\boldsymbol{T} \in \mathrm{R}^{n_a \times 9 \times 8}$. To generate the heatmap, we consider a rectangle area around the target vehicle that covers $y \in [-12\,\mathrm{m}, 75\,\mathrm{m}]$ and $x \in [-23\,\mathrm{m}, 23\,\mathrm{m}]$. In training, we set the spatial resolution to 1 m so the coordinate input has the shape $(46 \times 87, 2)$. During training and validation, the target vehicle is randomly selected among all vehicles that have complete 4 s records.

UQnet learns the spatial distribution of the last position after 3 s. We can use different sampling strategies to get predicted positions from the heatmap. The models on the leader board are ranked by Missing Rate (MR) so here we use a naive local-maximum sampling strategy (Gilles et al., 2021) to greedily generate the most possible k final positions ($k = 6$ for the INTERPRET challenge). MR is calculated as follows. If the predicted final position of the target agent is out of a given lateral or longitudinal area of the ground truth, it is "missed". MR measures how many per cent of predictions are missed. The lateral threshold is 1 m and the longitudinal threshold is a piece-wise function depending on the velocity of the target agent at the current moment:

$$\mathrm{th}(v) = \begin{cases} 1 & v < 1.4\,\mathrm{m\,s^{-1}} \\ 1 + \frac{v-1.4}{11-1.4} & 1.4\,\mathrm{m\,s^{-1}} \le v \le 11\,\mathrm{m\,s^{-1}} \\ 2 & v > 11\,\mathrm{m\,s^{-1}} \end{cases}$$

Besides MR, minFDE and minADE are also used to evaluate the quality of the predicted trajectories. minFDE is the L2-squared distance between the predicted endpoints and the ground truth, and minADE measures the average L2-squared distance between the forecasted trajectory and the label.

An MLP with 1 hidden layer completes the trajectory from the predicted last position. For uncertainty quantification, according to the empirical suggestion given in Kendall and Gal (2017), 7 randomly-initialized UQnets are trained in parallel (5–10 models are proper choices). Then the method proposed in Section 3.1 is used to quantify epistemic and aleatoric uncertainty.

Table 1 compares the performances of UQnet with other models on the leaderboard. Here UQnet is the performance of the first model (not the best one). UQnet (predictor only) does not use the causal regularization term. We see that the generalizability MR of UQnet is significantly improved from the previous-best 11.07% to 6.86% and the overall MR is reduced from 4.91% to 3.64%. UQnet's regular MR reaches the average level. If the causal regularization is not used, the regular MR is slightly better but the generalizability MR drops significantly.

Meanwhile, we also observe that the minFDE and minADE of UQnet are relatively worse than the state-of-the-art models. This is restricted by the used MR minimization sampling strategy, the resolution of the heatmap, and the trajectory generation method. The spatial resolution of UQnet is 0.5 m, which means it cannot give more accurate predictions within this mesh grid. The generated trajectory also ignores the kinematic constraints and the road layouts. Therefore, minFDE and minADE are degraded. This part can be technically improved but it is not relevant to the objective of this paper.

Next, we will focus on analysing the estimated uncertainty.

5.1. Precision–recall analysis

We first show the precision–recall curve for both aleatoric and epistemic uncertainty on the validation set. Here the precision is represented by hitting rate HR (HR = 1-MR) and log-likelihood (LL). We only sample 2 positions from the heatmap to calculate HR_2%. Fig. 6 shows how precision improves by preserving those cases with aleatoric or epistemic uncertainty lower than a specific threshold. For example, in Fig. 6(a), a point on the aleatoric uncertainty curve (the red line) at the 0.2 percentile depicts the hitting rate, considering the samples with the lowest 20% aleatoric uncertainty *only*. We see that the estimated uncertainty is negatively

Table 1
Comparison with other models.

Model	Regular			Generalizability			Overall		
Metrics	minADE	minFDE	MR	minADE	minFDE	MR	minADE	minFDE	MR
UQnet	0.3919	0.6215	1.96	0.6785	0.8608	**6.86**	0.4903	0.7037	**3.64**
UQnet-(predictor only)	0.3906	0.6190	1.84	0.7163	0.9150	10.31	0.5025	0.7206	4.75
GoHOME (Gilles et al., 2021)	0.1617	0.4578	**1.37**	**0.2747**	0.8685	11.68	0.2005	0.5988	4.91
Multimodal-Transformer	0.1417	0.4024	2.00	0.3494	0.8356	11.07	0.2130	0.5511	5.11
HDGT (Jia et al., 2022)	**0.1085**	**0.3361**	1.42	0.2806	**0.7484**	13.47	**0.1676**	**0.4775**	5.56
DenseTNT (Gu et al., 2021)	0.2819	0.6371	2.80	0.7256	1.0978	12.00	0.4342	0.7952	5.96

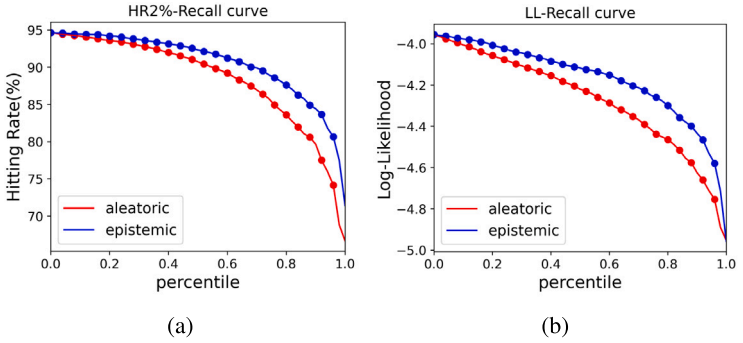


Fig. 6. Precision–Recall curve on the test set: (a) Hitting rate and (b) log-likelihood.

correlated with precision, which means the estimated uncertainty can indeed reflect the prediction confidence (Kendall and Gal, 2017).

Next, we will analyse aleatoric and epistemic uncertainty in detail. To facilitate the discussion below, now we split the validation set and the test set into the following three groups:

1 **Validation set**: All samples in the validation set are from exactly the same scenarios of the training set, including mergings, intersections, and roundabouts. It serves as a baseline.
2 **In-Distribution (ID) test cases**: Those test cases that are collected from the same scenarios as the training set and validation set.
3 **Out-of-Distribution (OOD) test cases**: Those test cases that are collected from new scenarios, including completely new types of scenarios, e.g. diverging.

5.2. Aleatoric uncertainty and predictability

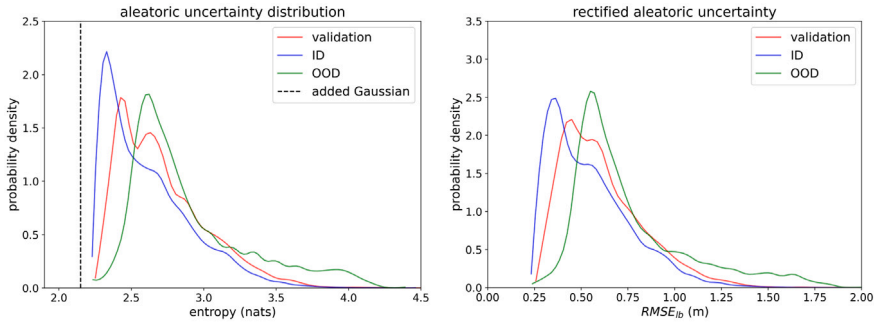
The distributions of estimated aleatoric uncertainty with added Gaussian noise for the three subgroups are shown in Fig. 7(a). They have similar shapes that are highly concentrated around 2.5 nats and the value is always higher than $H(\epsilon) = 2.12$ due to the added noise. We must rectify this estimation error. Unfortunately, although the added noise ϵ is independent of the prediction, differential entropy is not linearly additive. Note the rectified random variable of position as $\hat{\boldsymbol{Y}}_r = \hat{\boldsymbol{Y}} - \epsilon$, then $H(\hat{\boldsymbol{Y}}_r) \neq H(\hat{\boldsymbol{Y}}) - H(\epsilon)$. We must find a roundabout to correct this. Aleatoric uncertainty is closely related to the concept of "limit of predictability", which can be represented by the lower bound of accuracy for any model. In our previous study we show that the aleatoric uncertainty measured by conditional entropy gives the lower bound of negative-log-likelihood (NLL) and from that, we can further derive the limit of MSE (Li et al., 2022):

$$\mathrm{MSE}^2 = \det \boldsymbol{\Sigma} \geq \frac{1}{(2\pi e)^2} e^{2H} \tag{15}$$

MSE is linear and additive for independent random variables. So we can use (15) to derive the lower bound of MSE for the rectified prediction:

$$\mathrm{MSE}_{lb}(\hat{\boldsymbol{Y}}_r) = \mathrm{MSE}_{lb}(\hat{\boldsymbol{Y}}) - \sigma_\epsilon^2 = \frac{e^{H(\hat{\boldsymbol{Y}})}}{2\pi e} - 0.49 \tag{16}$$

Now we can use the formula above to convert entropy measures to MSE_{lb} (noted as σ_{lb}^2), or its square root RMSE_{lb}, to represent rectified aleatoric uncertainty equivalently. The results are shown in Fig. 7(b). The peak is now located at around 0.5 m. It is necessary to clarify that the σ_{lb}^2 is different from prediction MSE. For the two cases shown in Fig. 1, they have the same σ_{lb}^2 derived from the conditional entropy but (b)'s prediction MSE is apparently higher.



(a) (b)

Fig. 7. Distributions of aleatoric uncertainty for three different groups. (a) Differential entropy metric with Gaussian noise; (b) Denoised rectified RMSE lower bound metric.

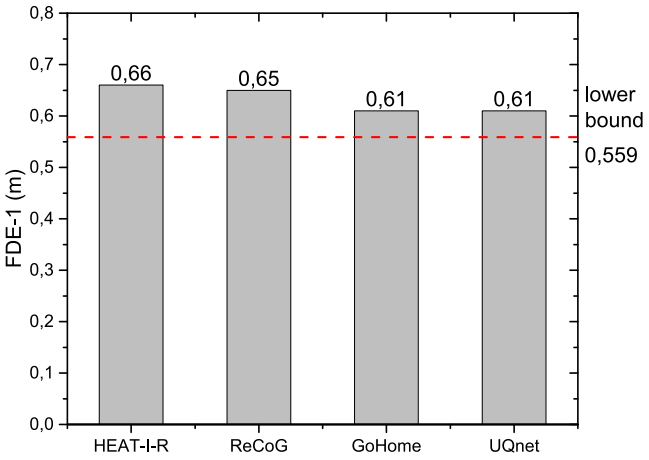


Fig. 8. The estimated minimum FDE for one prediction and the measured FDEs of HEAT-I-R (Mo et al., 2021), ReCoG (Mo et al., 2020) and GoHome (Gilles et al., 2021). All models are evaluated on the validation set.

In most trajectory forecasting tasks, people prefer to use final-displacement-error (FDE), which is the Euclidean distance. For given σ_{lb}^2, 2D symmetric Laplace distribution minimizes the L2 norm (Eltoft et al., 2006) (for 1 sampled prediction), noted as $p_L(x, y|\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \sigma_{lb}^2 \boldsymbol{I})$. However, the multivariate Laplace distribution contains the modified Bessel function of the second kind. Its mean-absolute-deviation does not have a closed form. So we can only estimate the lower bound of FDE by numerically calculating the following integral:

$$\mathrm{FDE}_{lb} = \int \sqrt{x^2 + y^2} \cdot p_L(x, y|\mathbf{0}, \sigma_{lb}^2 \boldsymbol{I}) dx dy \tag{17}$$

This estimated lower bound of average FDE (of one sampled final position) on the validation set is compared with the performances of some models in Fig. 8. It measures the limit of predictability from the perspective of probabilistic prediction. However, we emphasize that the analysis of predictability is only for a single prediction. In practice, we generally sample multiple predictions for safer motion planning. This strategy can significantly increase the coverage of outputs.

Fig. 9 shows three cases that have low epistemic uncertainty but high aleatoric uncertainty. Recall that low epistemic uncertainty means that the predicted heatmap is reliable and these cases are common. They represent three different types of missing information. Case-(a) has strong bi-modality due to the lack of turn signal. In the real world, the driver is expected to flash the left turn signal or do nothing. One modality would disappear. The high aleatoric uncertainty is caused by the special bird-view of the drone. In case-(b), the target vehicle is predicted to do a U-turn following a large vehicle. The heterogeneity of this steering manoeuvre itself is high. It highly depends on the driver's unknown proficiency. In case-(c), the target vehicle is driving at a high speed ($79.2\,\mathrm{km\,h^{-1}}$) and the leading car on its adjacent lane is significantly slower ($46.8\,\mathrm{km\,h^{-1}}$). Their gross longitudinal distance headway is about 21.5 m. We expect that the target vehicle will approach or even overtake the left neighbour in the next 3 s. The interaction uncertainty is high. The same acceleration will always yield the same speed difference. However, the higher the initial speed is, the bigger the difference in position becomes. The aleatoric uncertainty of cases-(a) can be reduced by adding turn signals, but adding sensors cannot help case-(b) and (c).

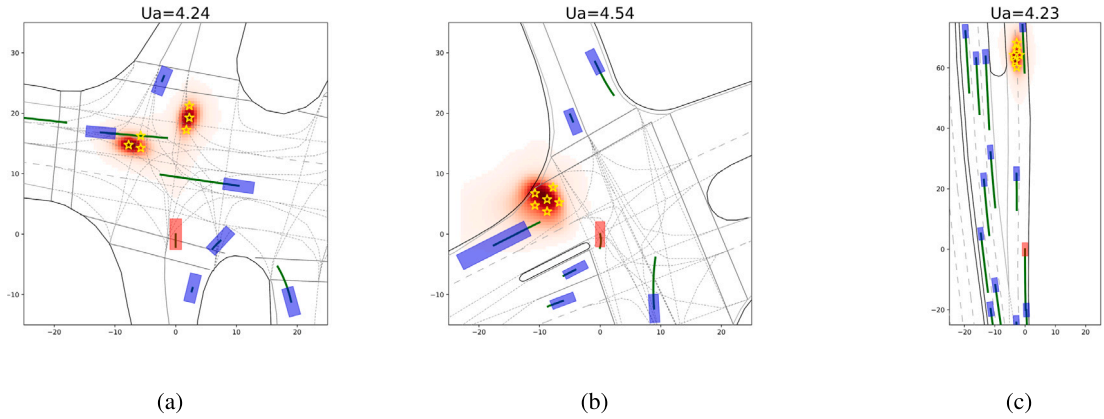


Fig. 9. Three examples of low epistemic uncertainty but high aleatoric uncertainty. The red vehicle is the target vehicle and blue vehicles are surrounding vehicles. Green lines are their trajectories in the past 1 s. The heatmaps represent the spatial distribution of the target vehicle's position after 3 s. Yellow stars are the 6 most possible sampled positions. U_a gives aleatoric uncertainty value (nats).

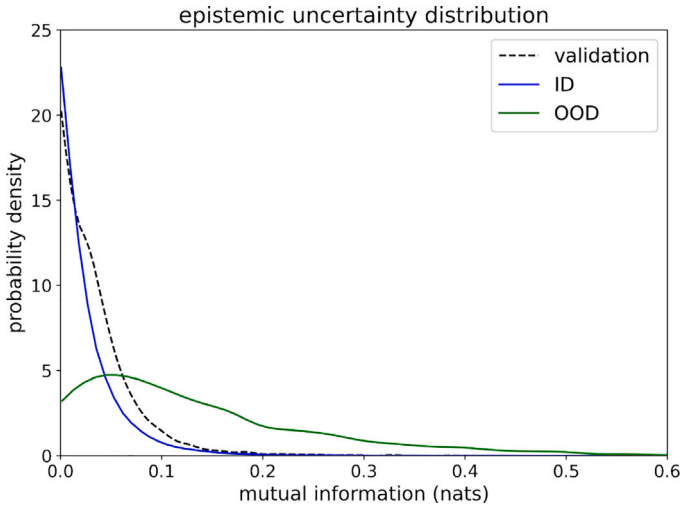


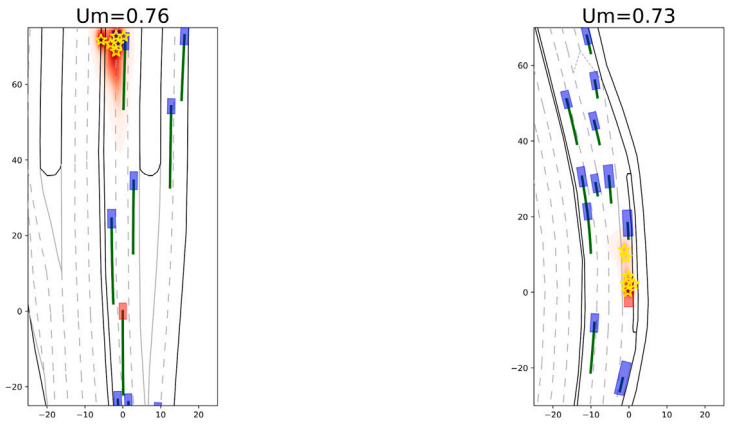
Fig. 10. The distributions of epistemic uncertainty.

5.3. Epistemic uncertainty and rareness

Fig. 10 presents the distributions of epistemic uncertainty. Different from Fig. 7, OOD's distribution is significantly different from the other two groups. The proposed method identifies more samples from new diverging/off-ramp scenarios as "new cases". But even in OOD, most samples are located at the low-epistemic end.

We found that most of these "rare cases" are from one specific scenario: the target vehicle is approaching the diverging decision point at a very high speed. Fig. 11(a) shows one example. We see that the spatial distribution disperses in a larger area because different models in the ensemble give inconsistent predictions. Some predictions are even off-road. The result is reasonable because the training set does not contain any diverging/off-ramp scenario. Although our model can effectively adapt to same-type scenarios, it cannot be generalized to a completely new situation. Fig. 11(b) shows another rare case. The target vehicle stops on the right-most lane but its leading and following vehicles are still moving. It is reasonable to infer that this is a rare abnormal case.

One may argue that, since intersections and roundabouts are composed of similar diverging and merging sub-structures, the model might also be applied to (on/off) ramps. However, the difference in operational speeds prohibits this generalization. When the speed is low (congested), epistemic uncertainty is small because the surrounding vehicles' influences are more important than geometry (the map elements). The target vehicle's choices are largely restricted. However, when the speed is high, driving behaviours are completely different. Human drivers will tend to keep longer distance headways and make faster and more determined decisions to avoid collision (Toledo et al., 2009) because the braking distance is proportional to the square of speed. Fig. 12 directly shows the relationship between the speed and epistemic uncertainty. The validation set and ID groups have very similar speed distributions (like the training set). Most of the samples are low-speed cases. There are not many cases where the target vehicle's speed is



(a) (b)

Fig. 11. Two examples of high epistemic cases. U_m is the estimated epistemic uncertainty (nats).

higher than $10\,\mathrm{m\,s^{-1}}$. The Pearson correlation coefficients are low (0.253 and 0.289). While the average speed of the OOD group is significantly higher. Speed and epistemic uncertainty are apparently positively correlated for OOD (0.711). Especially, those cases with speed higher than $20\,\mathrm{m\,s^{-1}}$ have significantly higher epistemic uncertainty. INTERPRET challenge splits the training set, the in-distribution test set, and the generalizability test set by scenario types, but the difference in speed is ignored and not balanced. However, pure data-driven models are not likely to be generalizable from low-speed to high-speed cases because their driving behaviours are substantially different. For example, the experiments in Huang et al. (2018) show that the intra-driver heterogeneity of car-following behaviours under high-speed situations is significantly higher than in low-speed situations. When the speed is high, even the same driver shows very different behaviours in keeping distance headways in repeated runs. Adding such extra domain knowledge established by traffic researchers is the key to further improving the generalizability.

Below we summarize the major findings in this section as follows:

- UQnet has good prediction performance. Compared with other correlation-based models, it has better generalizability due to the added causal regularization.
- The proposed heatmap-based UQ method gives estimates of both aleatoric and epistemic uncertainties.
- In this context,
  - Aleatoric uncertainty can be understood as a (reasonable) limit of predictability for any motion forecasting model.
  - Epistemic uncertainty is representative of new or rare interaction cases in the test set.
- Our results demonstrate that one of the major obstacles of domain generalization for motion forecasting is properly encoding the speed-dependency of the (causal) relationships between variables.

6. Conclusion and perspective

In this paper, we proposed a novel non-parametric spatial uncertainty quantification method. UQnet ensembles can give accurate predictions and reasonable measures of aleatoric and epistemic uncertainty. From aleatoric uncertainty, we estimate the lower bound of final-displacement error, which can measure the limit of predictability for trajectories. On the other hand, epistemic uncertainty can quantitatively identify which cases have not been seen in training. The main difficulty in improving models' generalization capabilities lies in improved modelling speed-dependent driving behaviours.

We offer two more tentative conclusions. First, we observed that aleatoric uncertainty is inherently high in many situations and our results seem to suggest that current AI models already harness the maximum value in commonly available trajectory data sets. In other words, using trajectory data and maps information only, more model sophistication is not likely to significantly improve the prediction performance for in-distribution situations. Combining and fusing more (different) data sources offers a more promising path to better predictions. For example, adding turn signals (blinkers) may effectively reduce much of the (aleatoric) interaction uncertainty. Second, our results support the idea that data-driven models should be combined with extra domain knowledge to gain better generalizability. A key example is the notion that operational speed matters for which of the available data (e.g. geometry versus surrounding vehicle kinematics) are most informative for motion prediction, and thus for the uncertainty associated with that prediction.

Finally, an important avenue for further research is the idea that epistemic uncertainty could be used as a tool to recognize rare events in training sets. Instead of collecting as much data as possible to improve generalization, one could rather focus on collecting

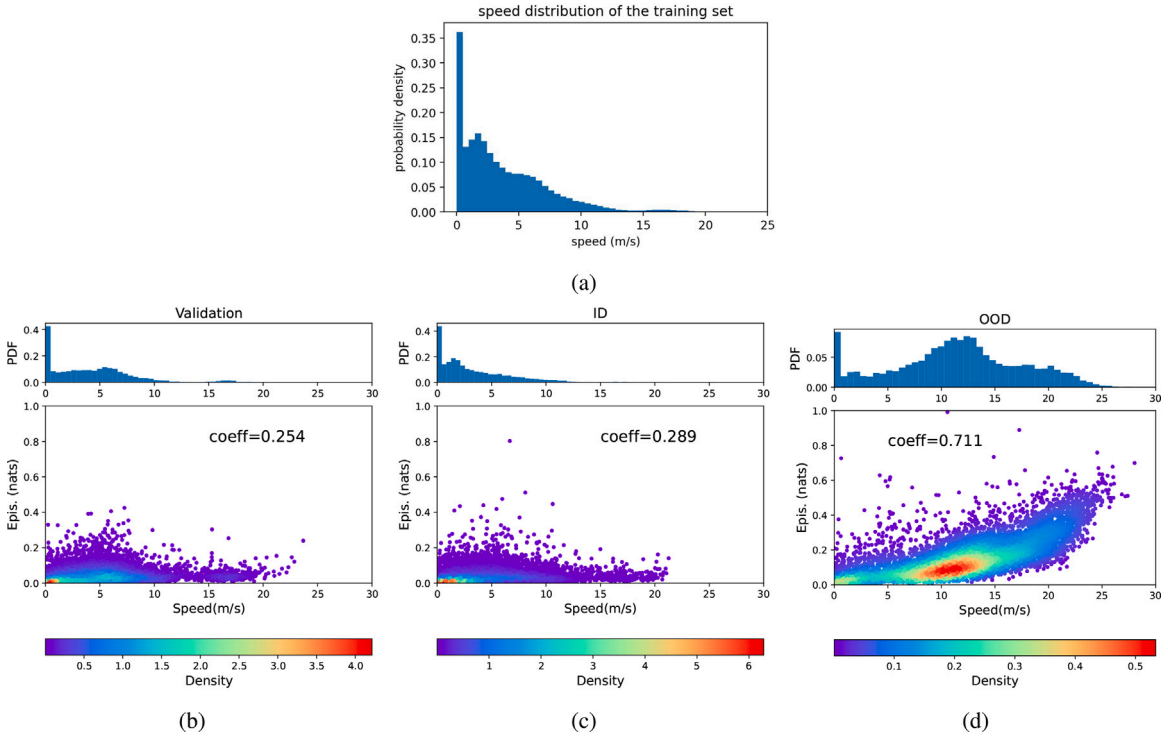


Fig. 12. (a) shows the distribution of the target vehicle's speed in the training set. (b) (c) and (d) show both the speed distribution and the scatter density plots of speed-epistemic uncertainty relationships for different groups. The Pearson correlation coefficient is also given.

sufficiently representative and heterogeneous data sets, in combination with applying adaptive learning techniques. This enables models to exploit continuous and guided online learning.

CRediT authorship contribution statement

Guopeng Li: Conceptualisation, Methodology, Software, Writing – original draft. **Zirui Li:** Methodology, Software, Writing – original draft. **Victor L. Knoop:** Methodology, Supervision, Writing – review & editing. **Hans van Lint:** Resources, Supervision, Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgements

This research is sponsored by the Dutch Research Council, Applied and Engineerign Sciences (NWO/TTW) project MiRRORS with grant agreement number 16270. We thank them for supporting this study.

Appendix A. Explanation of the causal regularization

Assume two random variables X and Y, Bayes' theorem says:

$$p_{Y|X}(y|x)p_X(x) = p_{X|Y}(x|y)p_Y(y) \tag{A.1}$$

However, the left and the right sides have different meanings from the perspective of causality. For instance, if X causes Y (denoted as $X \rightarrow Y$), then only the left side can be used. It means that we can control Y by manipulating X. The right side will fail when independent identical distribution does not hold (For example, eating more ice cream cannot elevate the temperature).

For the trajectory prediction and uncertainty quantification problem discussed in Section 3.2, in general, a model aims to build the following conditional probability distribution:

$$p(\boldsymbol{Y}|\boldsymbol{S}, \boldsymbol{M}, \boldsymbol{E}) \tag{A.2}$$

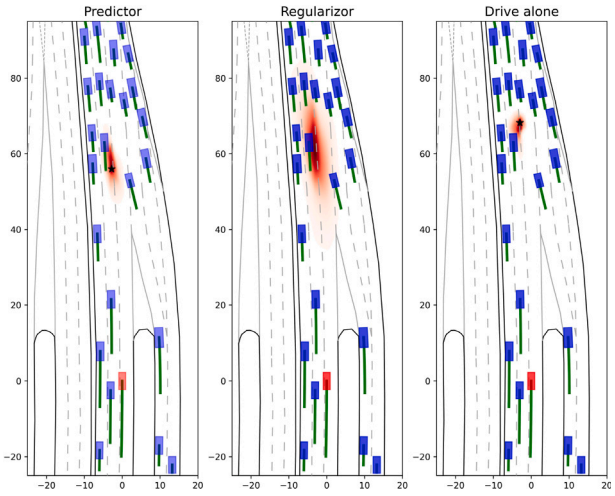


Fig. A.13. From left to right: output of the predictor $\hat{\boldsymbol{Y}}$, of the regularization net $\widetilde{\boldsymbol{Y}}$, and an imaginary case where all surrounding vehicles are removed from the inputs (drive alone). The black stars mark the most possible location after 3 s.

A correlation-based model (including most deep neural networks), directly learns this conditional distribution from (large) data but does not consider the causal relationships, which may lead to unsatisfactory generalizability and thus damage the estimate of epistemic uncertainty. Among the 3 input variables, $\boldsymbol{E}$ is absolutely needed. The problem is the positions of $\boldsymbol{S}$ and $\boldsymbol{M}$. We do know that $\boldsymbol{M}$ causes $\boldsymbol{S}$, not inversely. No matter how surrounding vehicles drive, maps are just there. However, changing maps will change drivers' intentions and behaviours. For Bayes formula:

$$p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{M},\boldsymbol{E}) = \underbrace{\frac{p(\boldsymbol{Y},\boldsymbol{S}|\boldsymbol{M},\boldsymbol{E})}{p(\boldsymbol{S}|\boldsymbol{M},\boldsymbol{E})}}_{\textbf{correct}} = \underbrace{\frac{p(\boldsymbol{Y},\boldsymbol{M}|\boldsymbol{S},\boldsymbol{E})}{p(\boldsymbol{M}|\boldsymbol{S},\boldsymbol{E})}}_{\text{wrong}} \tag{A.3}$$

Therefore, a naive but important rule for designing uncertainty-aware motion prediction models is:

The information can only flow from $\boldsymbol{M}$, and agents should not pass their information to the lanes.

From this perspective, the output of the normal predictor $\hat{\boldsymbol{Y}}$ and the regularization net $\widetilde{\boldsymbol{Y}}$ are defined as follows:

$$\begin{aligned}\hat{\boldsymbol{Y}} &\sim p(\boldsymbol{Y}|\boldsymbol{M},\boldsymbol{E},[\boldsymbol{S}|\boldsymbol{M},\boldsymbol{E}])\\ \widetilde{\boldsymbol{Y}} &\sim p(\boldsymbol{Y}|\boldsymbol{M},\boldsymbol{E}) = \int p(\boldsymbol{Y}|\boldsymbol{S})p(\boldsymbol{S}|\boldsymbol{M},\boldsymbol{E})d\boldsymbol{S}\end{aligned} \tag{A.4}$$

Notice that $\widetilde{\boldsymbol{Y}}$ does not exclude the influence of $\boldsymbol{S}$ but marginalizes it. To better show the differences, we choose a case from the test set (No. 1791) presented in Fig. A.13. The target red vehicle is approaching the merging at a high speed and there is congestion ahead of it. For the regularization net, because it considers all possible situations with diverse surrounding vehicles, the output covers a larger area, including constant speed cases, and acceleration/deceleration cases. The predictor's output is within it and, as expected, the vehicle will decelerate due to the congestion, and change to its left lane. As for the most right case, we construct an imaginary scenario in which all surrounding vehicles do not exist for the predictor. We see that the final location is about 13 m farther than the left case. The vehicle will simply follow the lane and keep a constant speed. This example shows that the influence (constraints) posed by surrounding vehicles is correctly learnt.

Fig. A.14 further shows the value of the regularization term (the leakage of probability) on the validation set and the out-of-domain test set. Most cases have low leakage due to the induced regularization term in training.

Appendix B. UQnet in detail

Many layers in UQnet are graph self-attention layers or cross-attention layers (Velickovic et al., 2017). We start from the more general cross-attention layer. Assume that we have a query input $\boldsymbol{X}_1 \in \mathrm{R}^{n_a \times d_a}$ and a feature input $\boldsymbol{X}_2 \in \mathrm{R}^{n_b \times d_b}$, and an adjacency matrix representing the connectivity $\boldsymbol{A} \in \mathrm{R}^{n_a \times n_b}$ then the output of a cross-attention layer is:

$$\boldsymbol{Q} = \boldsymbol{X}_1\boldsymbol{W}^q, \boldsymbol{K} = \boldsymbol{X}_2\boldsymbol{W}^k, \boldsymbol{V} = \boldsymbol{X}_2\boldsymbol{W}^v \tag{B.1}$$

$$\boldsymbol{H} = \mathrm{softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^T - 1e^7 \times (1-\boldsymbol{A})}{\sqrt{d_h}})\boldsymbol{V} \tag{B.2}$$

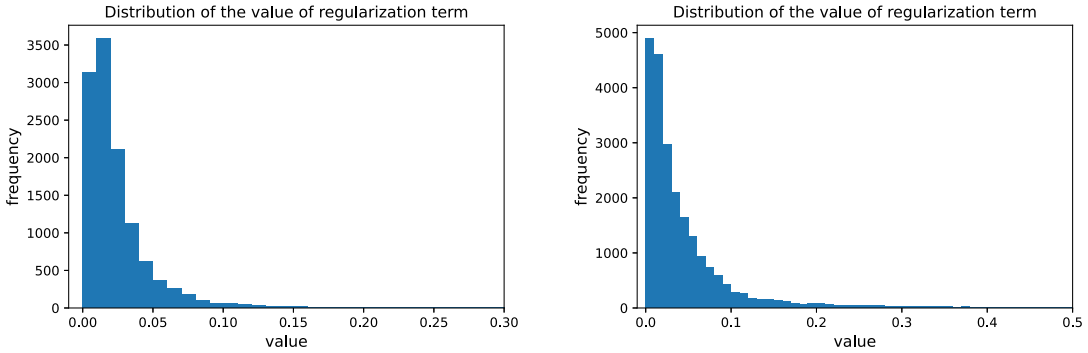


Fig. A.14. Distributions of the values of the regularization term on the validation set (left) and the out-of-domain test set (right).

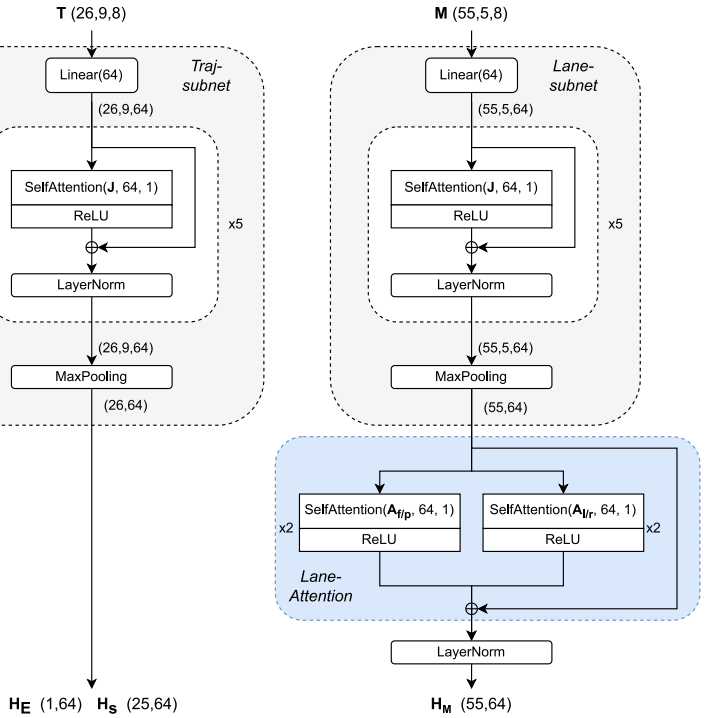


Fig. B.15. The encoder structure of UQnet.

where $\boldsymbol{W}^q \in \mathrm{R}^{d_a \times d_h}$, $\boldsymbol{W}^k, \boldsymbol{W}^v \in \mathrm{R}^{d_b \times d_h}$ are trainable parameters, the output is $\boldsymbol{H} \in \mathrm{R}^{d_a \times d_h}$. It can also be extended to multi-head attention layers by, for example, concatenating. We briefly note such a cross-attention layer with m attention heads as $\boldsymbol{H} = \mathrm{Attention}(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{A}, d_h, m)$. If $\boldsymbol{X}_1 = \boldsymbol{X}_2 = \boldsymbol{X}$, the cross-attention layer becomes a graph self-attention layer, briefly noted as $\boldsymbol{H} = \mathrm{SelfAttention}(\boldsymbol{X}, \boldsymbol{A}, d_h, m)$. The input of UQnet includes (the shape of each tensor is also given):

(1) $\boldsymbol{T}(26, 9, 8)$: trajectories of all agents in the selected rectangle. The first one is the target vehicle.
(2) $\boldsymbol{M}(55, 5, 8)$: map elements within the same range.
(3) $\boldsymbol{F}(N, 2)$: coordinates of a mesh grid. N depends on the resolution of the 2D histogram. It can be different during training and inference.
(4) $\boldsymbol{A}(81, 81)$: adjacency matrix that controls connectivity and information flow directions. It may vary in different layers according to the requirements. In this paper, we need the following adjacency matrices:

   (a) $\boldsymbol{J}$, matrix of ones.
   (b) $\boldsymbol{A}_a$, all agents are connected to all other agents and lanes, and lanes are connected bi-directional to each other.
   (c) $\boldsymbol{A}_{l/r/p/f}$, each lane only receives information from their left, right, previous, or following lanes and itself.
   (d) $\boldsymbol{A}_l$, all lanes are connected bi-directionally among them and only pass information to the target vehicle.

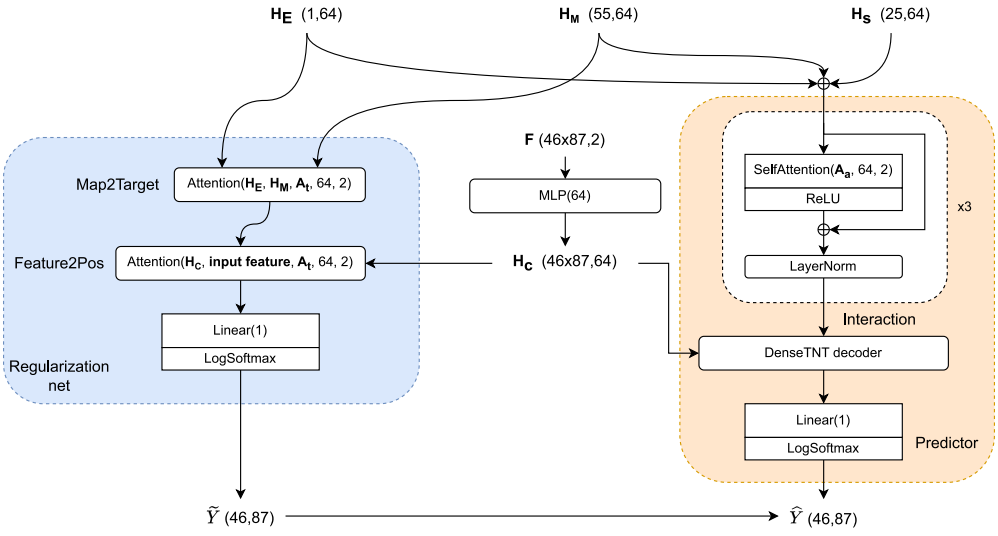


Fig. B.16. The decoder structure of UQnet.

The encoder and decoder structures are shown in Figs. B.15 and B.16. The *DenseTNT decoder* is given in DenseTNT (Gu et al., 2021). We refer the readers to their paper and the open-source code for more details.

References

Abdar, Moloud, Pourpanah, Farhad, Hussain, Sadiq, Rezazadegan, Dana, Liu, Li, Ghavamzadeh, Mohammad, Fieguth, Paul, Cao, Xiaochun, Khosravi, Abbas, Acharya, U Rajendra, et al., 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Inf. Fusion 76, 243–297.

Alahi, Alexandre, Goel, Kratarth, Ramanathan, Vignesh, Robicquet, Alexandre, Fei-Fei, Li, Savarese, Silvio, 2016. Social lstm: Human trajectory prediction in crowded spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 961–971.

Amini, Alexander, Schwarting, Wilko, Soleimany, Ava, Rus, Daniela, 2020. Deep evidential regression. Adv. Neural Inf. Process. Syst. 33, 14927–14937.

Ammoun, Samer, Nashashibi, Fawzi, 2009. Real time trajectory prediction for collision risk estimation between vehicles. In: 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing. IEEE, pp. 417–422.

Arnez, Fabio, Espinoza, Huascar, Radermacher, Ansgar, Terrier, François, 2020. A comparison of uncertainty estimation approaches in deep learning components for autonomous vehicle applications. arXiv preprint arXiv:2006.15172.

Bahari, Mohammadhossein, Saadatnejad, Saeed, Rahimi, Ahmad, Shaverdikondori, Mohammad, Shahidzadeh, Amir Hossein, Moosavi-Dezfooli, Seyed-Mohsen, Alahi, Alexandre, 2022. Vehicle trajectory prediction works, but not everywhere. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17123–17133.

Bergamini, Luca, Ye, Yawei, Scheel, Oliver, Chen, Long, Hu, Chih, Del Pero, Luca, Osiński, Błażej, Grimmett, Hugo, Ondruska, Peter, 2021. Simnet: Learning reactive self-driving simulations from real-world observations. In: 2021 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 5119–5125.

Caesar, Holger, Bankiti, Varun, Lang, Alex H., Vora, Sourabh, Liong, Venice Erin, Xu, Qiang, Krishnan, Anush, Pan, Yu, Baldan, Giancarlo, Beijbom, Oscar, 2020. Nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11621–11631.

Chang, Ming-Fang, Lambert, John, Sangkloy, Patsorn, Singh, Jagjeet, Bak, Slawomir, Hartnett, Andrew, Wang, De, Carr, Peter, Lucey, Simon, Ramanan, Deva, et al., 2019. Argoverse: 3D tracking and forecasting with rich maps. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8748–8757.

Chen, Guangyi, Li, Junlong, Lu, Jiwen, Zhou, Jie, 2021. Human trajectory prediction via counterfactual analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9824–9833.

Chen, Chacha, Wei, Hua, Xu, Nan, Zheng, Guanjie, Yang, Ming, Xiong, Yuanhao, Xu, Kai, Li, Zhenhui, 2020. Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (no. 04), pp. 3414–3421.

Cruz-Uribe, David, Neugebauer, C.J., 2002. Sharp error bounds for the trapezoidal rule and Simpson's rule. J. Inequal. Pure Appl. Math 3 (4), 1–22.

Daamen, Winnie, Loot, Martijn, Hoogendoorn, Serge P., 2010. Empirical analysis of merging behavior at freeway on-ramp. Transp. Res. Rec. 2188 (1), 108–118.

De Haan, Pim, Jayaraman, Dinesh, Levine, Sergey, 2019. Causal confusion in imitation learning. Adv. Neural Inf. Process. Syst. 32.

Der Kiureghian, Armen, Ditlevsen, Ove, 2009. Aleatory or epistemic? Does it matter? Struct. Saf. 31 (2), 105–112.

Djuric, Nemanja, Radosavljevic, Vladan, Cui, Henggang, Nguyen, Thi, Chou, Fang-Chieh, Lin, Tsung-Han, Singh, Nitin, Schneider, Jeff, 2020. Uncertainty-aware short-term motion prediction of traffic actors for autonomous driving. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2095–2104.

Ebrahimi, Sayna, Elhoseiny, Mohamed, Darrell, Trevor, Rohrbach, Marcus, 2019. Uncertainty-guided continual learning in Bayesian neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 75–78.

Eltoft, Torbjørn, Kim, Taesu, Lee, Te-Won, 2006. On the multivariate Laplace distribution. IEEE Signal Process. Lett. 13 (5), 300–303.

Etesami, Jalal, Geiger, Philipp, 2020. Causal transfer for imitation learning and decision making under sensor-shift. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (no. 06), pp. 10118–10125.

Fort, Stanislav, Hu, Huiyi, Lakshminarayanan, Balaji, 2019. Deep ensembles: A loss landscape perspective. arXiv preprint arXiv:1912.02757.

Gao, Jiyang, Sun, Chen, Zhao, Hang, Shen, Yi, Anguelov, Dragomir, Li, Congcong, Schmid, Cordelia, 2020. Vectornet: Encoding hd maps and agent dynamics from vectorized representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11525–11533.

Gilles, Thomas, Sabatini, Stefano, Tsishkou, Dzmitry, Stanciulescu, Bogdan, Moutarde, Fabien, 2021. Gohome: Graph-oriented heatmap output for future motion estimation. arXiv preprint arXiv:2109.01827.

Granger, Clive W.J., 1980. Testing for causality: A personal viewpoint. J. Econ. Dynam. Control 2, 329–352.

Gu, Junru, Sun, Chen, Zhao, Hang, 2021. Densetnt: End-to-end trajectory prediction from dense goal sets. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15303–15312.

Gupta, Agrim, Johnson, Justin, Fei-Fei, Li, Savarese, Silvio, Alahi, Alexandre, 2018. Social GAN: Socially acceptable trajectories with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2255–2264.

Hallé, Simon, Chaib-draa, Brahim, 2005. A collaborative driving system based on multiagent modelling and simulations. Transp. Res. C 13 (4), 320–345.

Houenou, Adam, Bonnifait, Philippe, Cherfaoui, Véronique, Yao, Wen, 2013. Vehicle trajectory prediction based on motion model and maneuver recognition. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 4363–4369.

Hu, Yeping, Jia, Xiaogang, Tomizuka, Masayoshi, Zhan, Wei, 2021. Causal-based time series domain generalization for vehicle intention prediction. In: NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications.

Huang, Yingfan, Bi, Huikun, Li, Zhaoxin, Mao, Tianlu, Wang, Zhaoqi, 2019. Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6272–6281.

Huang, Yanjun, Du, Jiatong, Yang, Ziru, Zhou, Zewei, Zhang, Lin, Chen, Hong, 2022. A survey on trajectory-prediction methods for autonomous driving. IEEE Trans. Intell. Veh..

Huang, Yong-Xian, Jiang, Rui, Zhang, HM, Hu, Mao-Bin, Tian, Jun-Fang, Jia, Bin, Gao, Zi-You, 2018. Experimental study and modeling of car-following behavior under high speed situation. Transp. Res. C 97, 194–215.

Jia, Xiaosong, Wu, Penghao, Chen, Li, Li, Hongyang, Liu, Yu, Yan, Junchi, 2022. HDGT: Heterogeneous driving graph transformer for multi-agent trajectory prediction via scene encoding. arXiv preprint arXiv:2205.09753.

Kendall, Alex, Gal, Yarin, 2017. What uncertainties do we need in bayesian deep learning for computer vision? In: Advances in Neural Information Processing Systems, vol. 30.

Kumor, Daniel, Zhang, Junzhe, Bareinboim, Elias, 2021. Sequential causal imitation learning with unobserved confounders. Adv. Neural Inf. Process. Syst. 34, 14669–14680.

Lakshminarayanan, Balaji, Pritzel, Alexander, Blundell, Charles, 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In: Advances in Neural Information Processing Systems, vol. 30.

Laxhammar, Rikard, Falkman, Göran, 2013. Online learning and sequential anomaly detection in trajectories. IEEE Trans. Pattern Anal. Mach. Intell. 36 (6), 1158–1173.

Leclercq, Ludovic, Knoop, Victor L, Marczak, Florian, Hoogendoorn, Serge P., 2016. Capacity drops at merges: New analytical investigations. Transp. Res. C 62, 171–181.

Lefèvre, Stéphanie, Vasquez, Dizan, Laugier, Christian, 2014. A survey on motion prediction and risk assessment for intelligent vehicles. ROBOMECH J. 1 (1), 1–14.

Li, Guopeng, Knoop, Victor L., van Lint, Hans, 2022. Estimate the limit of predictability in short-term traffic forecasting: An entropy-based approach. Transp. Res. C 138, 103607.

Liang, Ming, Yang, Bin, Hu, Rui, Chen, Yun, Liao, Renjie, Feng, Song, Urtasun, Raquel, 2020. Learning lane graph representations for motion forecasting. In: European Conference on Computer Vision. Springer, pp. 541–556.

Lin, Tsung-Yi, Goyal, Priya, Girshick, Ross, He, Kaiming, Dollár, Piotr, 2017. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988.

Liu, Yuejiang, Cadei, Riccardo, Schweizer, Jonas, Bahmani, Sherwin, Alahi, Alexandre, 2022. Towards robust and adaptive motion forecasting: A causal representation perspective. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17081–17092.

Lutteken, Niklas, Zimmermann, Markus, Bengler, Klaus J., 2016. Using gamification to motivate human cooperation in a lane-change scenario. In: Proceedings of the IEEE 19th International Conference on Intelligent Transportation Systems. ITSC, In: Series Using gamification to motivate human cooperation in a lane-change scenario, Rio de Janiero, Brasil, 2016 of Conference.

Ma, Yuexin, Zhu, Xinge, Zhang, Sibo, Yang, Ruigang, Wang, Wenping, Manocha, Dinesh, 2019. Trafficpredict: Trajectory prediction for heterogeneous traffic-agents. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, (no. 01), pp. 6120–6127.

Makansi, Osama, Ilg, Eddy, Cicek, Ozgun, Brox, Thomas, 2019. Overcoming limitations of mixture density networks: A sampling and fitting framework for multimodal future prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7144–7153.

Makansi, Osama, von Kügelgen, Julius, Locatello, Francesco, Gehler, Peter, Janzing, Dominik, Brox, Thomas, Schölkopf, Bernhard, 2021. You mostly walk alone: Analyzing feature attribution in trajectory prediction. arXiv preprint arXiv:2110.05304.

Malinin, Andrey, Gales, Mark, 2018. Predictive uncertainty estimation via prior networks. In: Advances in Neural Information Processing Systems, vol. 31.

Mo, Xiaoyu, Xing, Yang, Lv, Chen, 2020. Recog: A deep learning framework with heterogeneous graph for interaction-aware trajectory prediction. arXiv preprint arXiv:2012.05032.

Mo, Xiaoyu, Xing, Yang, Lv, Chen, 2021. Heterogeneous edge-enhanced graph attention network for multi-agent trajectory prediction. arXiv preprint arXiv:2106.07161.

Pang, Yutian, Zhao, Xinyu, Yan, Hao, Liu, Yongming, 2021. Data-driven trajectory prediction with weather uncertainties: A Bayesian deep learning approach. Transp. Res. C 130, 103326.

Poggi, Matteo, Aleotti, Filippo, Tosi, Fabio, Mattoccia, Stefano, 2020. On the uncertainty of self-supervised monocular depth estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3227–3237.

Prevost, Carole G., Desbiens, Andre, Gagnon, Eric, 2007. Extended Kalman filter for state estimation and trajectory prediction of a moving object detected by an unmanned aerial vehicle. In: 2007 American Control Conference. IEEE, pp. 1805–1810.

Rudenko, Andrey, Palmieri, Luigi, Herman, Michael, Kitani, Kris M., Gavrila, Dariu M., Arras, Kai O., 2020. Human motion trajectory prediction: A survey. Int. J. Robot. Res. 39 (8), 895–935.

Salzmann, Tim, Ivanovic, Boris, Chakravarty, Punarjay, Pavone, Marco, 2020. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In: European Conference on Computer Vision. Springer, pp. 683–700.

Sun, Pei, Kretzschmar, Henrik, Dotiwalla, Xerxes, Chouard, Aurelien, Patnaik, Vijaysai, Tsui, Paul, Guo, James, Zhou, Yin, Chai, Yuning, Caine, Benjamin, et al., 2020. Scalability in perception for autonomous driving: Waymo open dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2446–2454.

Tang, Xiaolin, Yang, Kai, Wang, Hong, Wu, Jiahang, Qin, Yechen, Yu, Wenhao, Cao, Dongpu, 2022. Prediction-uncertainty-aware decision-making for autonomous vehicles. IEEE Trans. Intell. Veh..

Toledo, Tomer, Koutsopoulos, Haris N., Ben-Akiva, Moshe, 2009. Estimation of an integrated driving behavior model. Transp. Res. C 17 (4), 365–380.

Velickovic, Petar, Cucurull, Guillem, Casanova, Arantxa, Romero, Adriana, Lio, Pietro, Bengio, Yoshua, 2017. Graph attention networks. Stat 1050, 20.

Vemula, Anirudh, Muelling, Katharina, Oh, Jean, 2018. Social attention: Modeling attention in human crowds. In: 2018 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 4601–4607.

Wang, Meng, Hoogendoorn, Serge P, Daamen, Winnie, van Arem, Bart, Happee, Riender, 2015. Game theoretic approach for predictive lane-changing and car-following control. Transp. Res. C 58, 73–92.

Wang, Hong, Lu, Bing, Li, Jun, Liu, Teng, Xing, Yang, Lv, Chen, Cao, Dongpu, Li, Jingxuan, Zhang, Jinwei, Hashemi, Ehsan, 2021. Risk assessment and mitigation in local path planning for autonomous vehicles with LSTM based predictive model. IEEE Trans. Autom. Sci. Eng..

Yoon, Seungje, Jeon, Hyeongseok, Kum, Dongsuk, 2019. Predictive cruise control using radial basis function network-based vehicle motion prediction and chance constrained model predictive control. IEEE Trans. Intell. Transp. Syst. 20 (10), 3832–3843.

Yuan, Kai, Laval, Jorge, Knoop, Victor L., Jiang, Rui, Hoogendoorn, Serge P., 2018. A geometric Brownian motion car-following model: Towards a better understanding of capacity drop. Transportmetrica B.

Zhan, Wei, Sun, Liting, Wang, Di, Shi, Haojie, Clausse, Aubrey, Naumann, Maximilian, Kummerle, Julius, Konigshof, Hendrik, Stiller, Christoph, de La Fortelle, Arnaud, et al., 2019. Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. arXiv preprint arXiv:1910.03088.

Zhao, Tianyang, Xu, Yifei, Monfort, Mathew, Choi, Wongun, Baker, Chris, Zhao, Yibiao, Wang, Yizhou, Wu, Ying Nian, 2019. Multi-agent tensor fusion for contextual trajectory prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12126–12134.

Zhou, Xingyi, Wang, Dequan, Krähenbühl, Philipp, 2019. Objects as points. arXiv preprint arXiv:1904.07850.